The assessment of liver fibrosis using the computerized analysis of ultrasonographic images. Is the virtual biopsy appearing as an option?

Monica Lupșor, Radu Badea, Sergiu Nedevschi, Cristian Vicaș, Simona Tripon, Horia Stefănescu, Corina Radu, Mircea Grigorescu

Abstract — Liver chronic diseases constitute an important public health issue. When diagnosing diffuse hepatopathies, ultrasonography is a simple method bringing useful information but not safe enough when determining the difference between certain diseases (steatosis, chronic hepatitis or early cirrhosis) or quantifying their severity. Present study is focused on examining pure fibrosis tissues in order to get a clear overview on how the presence of the fibrosis affects the ultrasonographic aspect of the liver. We tried to assess the usefulness of the computerized texture image analysis in noninvasive fibrosis grade quantification. From over 350 with biopsies we’ve selected 58 chronic hepatitis C patients, which have pure fibrosis without any steatosis. On each image we established a Region of Interest and we extracted 166 features using 4 algorithms from it. We compared the mean values of each feature between medical significant fibrosis stages using Student’s test (p<0.05). We find that we can distinguish between close fibrosis stage with one or more relevant features, but there is no feature that has a relevance over 95% in each comparison case. Therefore, using only one parameter cannot distinguish between all liver fibrosis stage, and a combination of features must be used in order to successfully diagnose the fibrosis stage.

Keywords: fibrosis, image processing, non-invasive, texture, ultrasonography

1. INTRODUCTION

Liver chronic diseases constitute an important public health issue. The infection with B, C or Delta hepatic viruses, the non-alcoholic fatty liver disease or the alcoholic hepatopathy represent the vast majority of liver diseases, other dysfunctions (genetic, metabolic or immunological) having a lower incidence.

The prevalence in Romania (32%), extrapolated for the general population, exceeds the European average (15-20%) concerning chronic hepatopathies. The evolution of diffuse liver diseases varies, but generally is quite long. Even the most severe chronic hepatopathies have a slow but progressive evolution, which lasts decades, often over 20-30 years. Whatever the nature of the liver aggression, it seems to follow the pattern: inflammationÆnecrosisÆhealing (fibrosis)Æregeneration (cirrhosis)ÆdysplasiaÆhepatocellular carcinoma.

Fibrosis is the scarring response formed in the chronic injury of any cause. It is a dynamic process, with a possibility of reversibility. For the moment, the golden standard in evaluating fibrosis is the liver biopsy. Using the liver biopsy one can establish with certainty the diagnosis, one can assess the severity of necroinflammation and fibrosis and one can evidence the simultaneous liver diseases. On the other hand, it is an invasive procedure, with possible side-effects (pain in 30-40% of the cases, hemorrhage, biliary peritonitis, penetration of abdominal vessels, pneumothorax - 3‰ or even death – 2 / 10.000 of the cases). [1],[2] Additionally, in 24% of the cases there can be sampling errors (either by the fragmentation of the biopsy specimen or by removing an inadequate volume). As a matter of fact, the fibrosis distribution inside the liver isn’t always homogenous and the biopsy specimen accounts, in average, for just 1/125,000 of the liver volume (in average it has the shape of a cylinder with a diameter of 1 mm and a length of 1,5 cm). One can also add the intra- and inter-operator variability in assessing fibrosis (reported in 10-20% of the cases). The two combined factors (the lack of representativity of the biopsy specimen and the variability in the assessment of fibrosis) lead to a cirrhosis false negative rate of 24%.

Therefore it is important to assess as correctly as possible the fibrosis in a non-invasive way, using biochemical and
imaging methods, as an alternative to the liver biopsy. Non-invasion is a principle worth to be applied in any situation, especially in chronic hepatopathies where the fibrosis quantification has an important prognostic value, considering that it allows the appreciation of the progression risk towards cirrhosis, playing an important part in: determining the therapeutical decision and appreciating the efficiency of anti-fibrotic treatment. [3] The imaging diagnosis methods have the advantage, besides non-invasion, of evaluating the entire organ, offering a more exact appreciation of the disease severity if the fibrosis does not uniformly affect the liver. The main methods used are: ultrasonography (US), computer tomography (CT), magnetic resonance imaging (MRI) and elastography (ultrasonographic or by magnetic resonance). Out of these, ultrasonography is the most used imaging method, taking into account, among others, the lack of harmful effects and the low cost.

2. LIVER FIBROSIS ASSESSMENT USING ULTRASONOGRAPHY

Usually, the imaging examination for the liver fibrosis assessment has been limited in detecting cirrhosis and its complications. The US, CT and MRI studies were based on the identification of morphopathologic modifications at the liver level during the natural evolution of cirrhosis, such as the diminution of the right lobe size and concomitantly, the increase of the left lobe and of the caudate lobe size. These approaches are characterized by a high specificity but they have a limited sensitivity considering the fact that the significant morphological modifications are present only in the advanced phases of the disease.

In concordance with the high clinical and biochemical variability of chronic hepatitis (one of the principal causes of fibrosis), the ultrasonic exam varies from the “normal” aspect to modifications similar to those of the liver cirrhosis. The most frequent changes are: hepatomegaly, slightly increased echogenity, sometimes with moderate attenuation (external toxic factor), homogenous structure, granular, or even non-homogeneous structure, but without clear focal images, regular capsular contour, discrete dilatation of portal venous system, splenomegaly, adenopathies in the hepatic hilum (in case of viral replication). [4] However, there are non-specific changes approximately specifying the etiology. They need to be correlated with the clinical and biological features and liver biopsy. Ultrasonography is useful in this situation in order to exclude (with probability) portal hypertension; it cannot exclude the incipient portal fibrosis. At the same time, the differential diagnosis with early cirrhosis is difficult to be determined and the patient needs an ultrasound reevaluation every 6-12 months.

When diagnosing diffuse hepatopathies, US is a simple method bringing useful information but not safe enough when determining the difference between certain diseases (steatosis, chronic hepatitis or early cirrhosis) or quantifying their severity. Although these pathological conditions are different (as substrata), the main obstacle when differentiating them is the extremely subtle “visual” differences shown on the US image.[5] The visual discrimination criteria depend on the subjective interpretation of the examiner which may lead to the limitation of the method’s reproducibility and diagnosis errors. This is the reason why the usual US examination attempts to be optimized.[6] One approach may be the computer processing data forming the US image, taking into consideration the fact that all information related to the tissular characters already exists in the echoes sent back to the transducer.

3. OPTIMIZING THE USUAL ULTRASONOGRAPHY EXAMINATION, USING THE COMPUTERIZED IMAGE PROCESSING.

This method is based on the principle according to which the pathological tissular modifications due to a specific disease (such as steatosis, chronic hepatitis with different fibrosis stages, or the early cirrhosis) determine alterations of the physical and micro architectural features (density, thickness, elasticity, homogeneity, etc.).These are very difficult to visualize, but because they affect the propagation of the ultrasounds, they can be perceived through the complex analysis of the image (the ultrasonic tissular characterization) as a different textural pattern as opposed to the healthy one. [7]

3.1. Motivation

This paper presents some preliminary results in the domain of ultrasonographic fibrosis evaluation and “virtual biopsy” concept. In the first steps of achieving these goals one need to find, describe and understand an imagistic model of fibrosis.

There are several ways to describe an image, one way to achieve this is by computerized texture evaluation. A texture can be described in terms of texture features. Each texture feature is computed with an algorithm and is a measure of various visual or non visual aspects of the texture. One problem that arises is what features we will use and what are the thresholds that will better discriminates between various grades of fibrosis.

Present study is focused on examining pure fibrosis tissues in order to get a clear overview on how the presence of the fibrosis affects the ultrasonographic aspect of liver. For this reason we avoid patients with associated steatosis even if the steatosis was negligible (below 10%). We established an examination protocol based on doctor’s experience and visual aspect of the image. A goal of this study is to evaluate this protocol and to change it if necessary.

3.2. Material and methods

We examined 350 patients suffering from different diffuse liver diseases, who have had a liver biopsy performed in the same day, for the quantification of
fibrosis stage. From over 350 with biopsies we’ve selected 58 chronic hepatitis C patients, which have pure fibrosis without any steatosis. We have also 6 patients that has no diffuse hepatopathies. (without any fibrosis or steatosis).

The fibrosis stage for patients suffering from C viral chronic hepatitis – selected for the study – has been quantified histopathologically by using the Metavir score system, as it follows:

- Stage 0 (F0) – no fibrosis
- Stage 1 (F1) – portal fibrosis without septa
- Stage 2 (F2) – portal fibrosis and few septa
- Stage 3 (F3) – numerous septa without cirrhosis
- Stage 4 (F4) – cirrhosis

The ultrasound exam was performed on a GE Logiq7 ultrasound machine, using a convex probe of 5.5 MHz. Ultrasonographic images were processed using a software tool developed at Technical University Cluj-Napoca.

The examination protocol was built in order to improve the quantity of the information gathered from the tissue and to lower the noise level. We tried to maximize the number of pixels and gray levels that are allocated to the liver tissue in the ultrasonographic image. We used harmonics examination, a higher value for dynamic range. We set 2 focal points. Their position (depth) was manually changed by the physician according to patient’s anthropometric characteristics in such a way that the focal points were situated 1-2 cm below the liver capsule. Time gain compensation curve was set into neutral position. We avoided using any post processing because it alters the texture information. We tried to maintain a good frame rate in order to avoid motion blur when we freeze the image. Images were acquired digitally on the ultrasonographic scanner. We used a lossless file format (BMP) in transferring images from ultrasonographic scanner to the computer.

Using these settings, from each patient we acquired following image categories:

- Depth set at 16 cm (in order to see an overview of the liver)
- Depth set at 8 cm (in order to have a better image over the liver texture)

This protocol was applied for the left lobe and right lobe. In figure 1 we have some images exemplifying the protocol. We also used the “Write Mode” setting in order to extract information from a limited area of the tissue (the area was established by the physician) by increasing the number of pixels per mm. (figure 2). We tried to capture images without artefacts and with as much as possible tissue.

In our study we’ve compared images from the left and right lobe. In total we evaluated 836 images. On each image from the right lobe we set a Region of Interest (ROI) of 64x64 pixels. On images from the left lobe a smaller ROI was set (32x32 pixels). The ROI was chosen such that it is situated right beneath the liver capsule with the centre as close as possible to the centre of the ultrasonographic image. The ROI has to avoid any artefacts like blood vessels, costal shadows, spikes etc. In the centre of the image and in the focal region we can find the best image quality (in terms of resolution and noise levels) [4] In figure 3 we show an example of ROI fixed on an image.

From each Region of Interest we’ve extracted 166 features using 4 algorithms using various parameters. The algorithms are the following:

- Modified differential Box Counting (MDBC)
  Fractal analysis is becoming more and more popular in image processing community. The texture feature that can be derived using fractal analysis is the fractal dimension. In literature it is believed that fractal dimension correlates with the judgment of roughness by human vision. Using the method proposed by [8] we calculate 15 features by varying the algorithm parameters.

- Gray Level Co-occurrence Matrix (GLCM) is a very common algorithm used in texture analysis.[9] We compute second order statistics on a specially constructed matrix. In early
studies it is shown that human vision hardly discriminates tow textures if their second order statistics are the very close. [10]. We compute statistics like: contrast, entropy, energy and inverse difference. We compute 120 features by varying the algorithm’s parameters (distance, angle).

- **First order histogram based features (FO).** We apply various formulas in order to characterize the shape of the ROI histogram. We compute mean, variance (the same meaning as in the statistics), skewness (the symmetry of the histogram), kurtosis (measures the flatness of the histogram), energy and entropy (measures the uniformity of the histogram) [10] We compute 6 features.

- **Gray Tone Difference Matrix (GTDM)** is another algorithm that tries to define texture features that correlate with human perception [10]. The features that are computed are: coarseness (defined by the size of texture primitives), contrast (dependent on the intensity difference between neighbouring pixels), busyness (described by high spatial frequency of intensity changes), complexity (dependent on the number of different primitives and different average intensities) and texture strength (clearly definable and visible primitives). Based on these features and varying the window size we computed 25 features.

After computing these 166 features for 836 images, we try to evaluate if there are differences between the values of a feature when we examine a level of fibrosis compared to another level of fibrosis. We focus on the medical significant differences (ex: F0-F1, F1-F2, F3-F4). We partitioned the images in groups by fibrosis diagnostic and by protocol. We compare two fibrosis grade in images taken by the same protocol.

For all the images in a group we compute the mean value of each texture feature. Student t test is used to assess the statistical relevance of the difference. We considered that a mean difference is relevant for a $p<0.05$. For each comparison we sort descending the features according to their relevance. An empirical measure of the “easiness of discrimination” is the number of the features that are relevant. If many features are relevant for a comparison we assume that a classifier will have very good performances (specificity and sensibility) in discerning between two groups.

After evaluating the relevance of each feature in four type of images (8 cm left lobe, 8 cm right lobe, 16 cm left lobe, 16 cm right lobe) in four comparisons (FO-F1, F1-F2, F2-F3, F3-F4) we tried to get a larger view over the behaviour of each feature. We grouped the features according to image type. In each image type lot we computed the geometrical mean between the discrimination relevance of each feature.

### 3.3. Results

In table 1 we show the distributions of ultrasonographic images relative to protocol and fibrosis grade.

<table>
<thead>
<tr>
<th>Fibrosis grade</th>
<th>Number of patients</th>
<th>Number of images from the right lobe</th>
<th>Number of images from the left lobe</th>
<th>Write Zoom</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>8 cm</td>
<td>16 cm</td>
<td>8 cm</td>
</tr>
<tr>
<td>0</td>
<td>6</td>
<td>30</td>
<td>32</td>
<td>12</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>42</td>
<td>49</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>49</td>
<td>56</td>
<td>36</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>59</td>
<td>72</td>
<td>27</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>39</td>
<td>46</td>
<td>23</td>
</tr>
</tbody>
</table>

#### 3.3.1. Comparison between fibrosis level 0 and fibrosis level 1

In Write zoom there are no relevant features ($p<0.05$) but in images taken at 16 cm from the right lobe we have 16 features that are relevant. We have the MDBC algorithm, some GTDM statistics (Coarseness and Texture strength) and two features from the FO algorithm (Skewness and Kurtosis).

At 8 cm/ right lobe we have only one relevant feature, FO Skewness.

From the left lobe, at 16 cm we have 15 relevant features. Again we find MDBC with 4 features, followed by 3 GTDM features (busyness). The rest of 8 features are a combination of GTDM, MDBC and FO.

At 8 cm, left lobe we find only one feature, Inverse Difference from GLCM.

#### 3.3.2. Comparison between fibrosis level 1 and fibrosis level 2

In Write Zoom there are no relevant features although MDBC gives us one 94% relevant feature.

In the 16 cm/right lobe we found 9 relevant features: GLCM (Contrast and Inverse Difference) and GTDM (Busyness). In the 8 cm/right lobe we find 56 relevant features. The most relevant features are given by MDBC (3 features) followed by 2 of GTDM (Texture strength). The rest of relevant features are given by Energy and Entropy calculated by GLCM algorithm using various distances and two FO features (Entropy and Energy).

From the left lobe at 16 cm we have 56 relevant features. At first 8 positions we have MDBC features followed by GLCM (entropy, energy), GTDM (Coarseness, and Complexity) and 2 FO features. At 8 cm we have 69 features. Again, 4 MDBC features followed by GTDM (Texture strength, Coarseness) and GLCM (Energy and Entropy).

#### 3.3.3. Comparison between fibrosis level 2 and fibrosis level 3

In Write Zoom protocol we have 72 relevant features. In the topmost position we can find GLCM with Contrast and Inverse difference at various distances ( 9 features) MDBC ( 1 feature) followed by various features from GTDM (Complexity , GLCM (energy, entropy, contrast)
again and MDBC. We notice that FO algorithm gives us one (variance) relevant feature.

At 16 cm/right lobe we find 109 relevant features. In first position we have one MDBC feature followed by 5 Texture Strength features from GTDM. The following positions are occupied by GLCM, GTDM and FO features.

At 8 cm/right lobe we have 143 relevant features. In the first position we have Energy from FO algorithm followed by 53 GLCM (energy and entropy) statistics. We also have MDBC, GTDM (Complexity, Busyness) and GLCM (Inverse Difference and contrast) features.

From the left lobe, at 16 cm we find 77 relevant features. Most of relevant features are from GTDM (Energy and entropy) followed by MDBC and GTDM (Texture strength and Coarseness). At 8 cm we find 58 relevant features. We find again GLCM statistics (energy and entropy) mixed with MDBC features. We find 5 GTDM features and one FO (Energy).

3.3.4. Comparison between fibrosis level 3 and fibrosis level 4

In Write Zoom protocol we have 4 relevant features all given by Complexity from GTDM.

In 16 cm/right lobe we find 6 relevant features: Mean from FO, one MDBC feature and Contrast from GLCM computed using 4 different distances.

In 8 cm/right lobe one can notice that there are 149 relevant features. In the first position we find again Mean from FO followed by Contrast (GTDM), Entropy from GLCM and one feature from MDBC. In the next position we can find Contrast, Entropy, Energy, Inverse Difference from GLCM mixed with Contrast, Texture Strength from GTDM and some MDBC features.

In the features extracted from the left lobe, at 16 cm we find 7 relevant features. 3 GLCM features (inverse difference) followed by 2 MDBC features and 2 GTDM features (Busyness).

At 8 cm we have 12 relevant features. GTDM (Texture Strength) and GLCM (contrast, Inverse difference) The first MDBC feature has a relevance of 94.5% (p=0.055)

3.3.5. Texture feature comparison

For each acquisition protocol we established an order of relevance. In table 2 we can see the best value of geometrical mean for each protocol. One can see that in left lobe at 16 cm we obtained the best geometrical mean relevance value.

<table>
<thead>
<tr>
<th>Table 2. Geometrical mean between relevance grade in each image acquisition protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquisition protocol</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Right lobe, 16 cm</td>
</tr>
<tr>
<td>Right lobe, 8 cm</td>
</tr>
<tr>
<td>Left lobe, 16 cm</td>
</tr>
<tr>
<td>Left lobe, 8 cm</td>
</tr>
</tbody>
</table>

3.4. Discussions

According to our proposed method of estimating the “easiness of discrimination” by counting the number of the features that are relevant in each case, we evaluate the findings in our comparisons.

We try to discover if there is a fibrosis pattern, if there is one (or more) relevant features for all comparisons that can be used to distinguish between the various stages of fibrosis.

Regarding the comparison between F0 and F1, the results from the left lobe are consistent with the findings from right lobe at the same depth in terms of number of relevant features. When we compare F1 with F2, we get better results from the left lobe at 16 cm. Comparing F2 and F3, in all combinations of depths and lobes GLCM statistics energy and entropy behaves very well. In the comparison between F3 and F4, all results are consistent except form those at 8 cm from right lobe where the number of relevant features is far greater (149 compared to ~10). This might indicate us that in evaluating cirrhosis, images from right lobe at 8 cm are the most relevant.

We can easily see that although we can distinguish between close fibrosis stages, every comparison has different relevant features, found in different lobes and there is no “common rule” that matches all. The fact that in each comparison case there is at least one relevant feature is the reason to believe that our examination protocol is good for our purpose. Examining the liver using only one protocol (i.e. Right lobe, x cm depth) is not enough to asses a correct fibrosis diagnostic. An overview of the liver tissue is necessary in order to assess a correct fibrosis stage diagnostic.

Next we are focusing on individual feature performances by computing the geometrical mean of relevance. One can notice that we have a good mean relevance only in left lobe at 16 cm. In the rest of the protocols we have at least 2 values with relevance lower than 95%. We decided to further investigate this protocol. In 16 cm left lobe, in the first 3 positions we find MDBC features with mean relevance over 95%. There is not an specific order in mean values. According to the literature, the roughness of the liver tissue should increase as the fibrosis grade increases.[9] Here we do not confirm this aspect. This can be due some errors (ROI positioning, acquiring errors, noise in the images), reduced volume of data (only 58 patients) etc. However our conclusion is that using only one parameter cannot distinguish between liver fibrosis stages. One can also notice that there is not a single feature that has relevance over 95% in all 4 comparison situations. These findings show that simple statistics are not enough to capture the differences between various fibrosis grades and encourages us to develop and use methods like classifiers. There is not enough to use only one feature to distinguish between various fibrosis stages. The results of statistical test Student is influenced by the volume of the studied lot. Some lots have a smaller volume that others (Write Zoom lot compared to the lot acquired at 16 cm from right lobe). Some results might get improved if we will add more patients to the study.

An important fact in the diagnosis of the diffuse liver diseases is to trace the difference between steatosis and other diseases, defined by the increase of the ecogenity. Steatosis and fibrosis in particular can have the same...
ultrasonic aspect. Many times they coexist, that is why the “fatty-fibrotic pattern” term is used to define the aspect resulted [11]. Although these pathological conditions are different (as substrata), the main obstacle in their differentiation is the extremely subtle “visual” differences that they generate on the ultrasonic image. There have been attempts in differentiating steatosis from fibrosis on the ultrasonic image. In clinical practice, fibrosis, in absence of fat may not be associated with increased attenuation. Certainly, one study concluded that fat alone accounted for the increased attenuation in patients with cirrhosis, although another study did suggest that, in vitro, fibrosis causes some attenuation, but only half that of fat. [11, 12] Fibrosis may also be distinguished from fat by the course echo pattern produced and the increased definition of portal vein. But, in the end, all these criteria are just visual and subjective. Therefore, further work has to be done in order to study the effect of steatosis in liver tissue over the texture features and study methods to differentiate between steatosis and fibrosis. In the future we want also to focus on the relevance of each feature by using the Principal Component Analysis algorithm. This will give us the relevant features in terms of liver texture description. These features will be used in implementing a classifier. Fractal based features like MDBC behaved stable across the studied lots. Although in some cases other features performed better (GLCM, energy and entropy) we believe that more algorithms should be implemented in order to estimate texture fractal dimension.

4. CONCLUSIONS

Even if the usual ultrasonographic examination cannot always clearly make the difference between different types of affections (as the chronic hepatitis, steatosis or incipient cirrhosis) or between their degree of severity, the results can be highly improved by using ultrasonic image processing methods. Our study shows that texture analysis is a promising way in differentiating the fibrosis stages. Algorithms like GLCM and MDBC gives us punctual relevant features. Although there is not a single feature that is relevant in all comparison cases (and this means that we cannot distinguish between various stages of fibrosis using only one feature), our findings suggest that a combination of features must be used in order to successfully diagnose the fibrosis stage. This finding is consistent with literature, where algorithms called classifiers are used in combination with texture features in order to assess the liver fibrosis stage.

The direct medical benefit of computerized methods of image analysis will resume in a possible early and more accurate diagnosis of diffuse liver diseases, avoiding as much as possible the detrimental effects of the invasive diagnosis methods. The immediate application of this method is the possibility of discerning between patients with cirrhosis (F4), patients with severe fibrosis (F ≥ 2) and patients with no (or moderate) fibrosis (F<2) using a noninvasive method. These preliminary results justify taking into consideration a larger group of patients in order to validate the method and implement in current clinical practice the concept of “virtual biopsy”.

5. REFERENCES